

Species Identification in Diatoms using DNA Barcoding: An Overview

Suvechha Kabiraj¹, Suman Jyoti Bhuyan¹ and Umesh Goutam^{1,2*}

¹Department of Molecular Biology and Genetic Engineering, School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, Punjab, India.

²Department of Biotechnology, School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, Punjab, India.

(Corresponding author: Umesh Goutam*)

(Received 28 June 2022, Accepted 04 August, 2022)

(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Diatoms are unicellular photoautotrophic microalgae found predominantly in freshwater and marine environments, but sometimes in soil and as aeroplankton. Diatoms are vital components of ecosystems as food sources for a variety of different creatures. Because many morphological features can only be identified using scanning electron microscopy or other high-resolution technologies, identifying diatoms morphologically below the genus level requires specialist taxonomic knowledge and often expensive infrastructure. Alternatives include DNA barcoding and high-throughput-sequencing which allows for the quick analysis of a large number of collected samples at a lower cost than microscopy. Therefore, in order to appropriately identify environmental sequences, a carefully managed reference library is required. Standardised processes now rely on microscopic measurements, which are gradually takes time and highly susceptible to misidentification. To address these issues, DNA barcoding is better alternative. A large number of barcodes can be captured from natural materials when barcoding is applied in combination with next-generation sequencing. By analysing the sequences to a reference genomic library and employing algorithms, these barcodes are classified as specific diatom taxa. The application of the DNA barcoding idea to diatoms has a lot of promise for resolving the problem of erroneous species identification and thereby facilitating biodiversity assessments of environmental samples. DNA barcodes in diatoms can be used for a variety of applications, including classification of taxonomic group using DNA, determining genetic variation in a specific circumstance. Researchers are currently interested in developing DNA barcodes for all living organisms and compiling data that will be available to the public to aid in the understanding of the world's natural biodiversity. The identification of unidentifiable biological material to a taxonomic group and species diversity of living organisms can be done using accurate and reliable information about DNA barcoding. The challenges include while performing this study are the phylogenetic framework of barcoding, development and testing of candidate barcodes and creation of diatoms and the emergence of a system of identification. Important future challenges will also focus on building a DNA barcode library and developing genomic sequencing methods as efficient as possible by utilizing these genetic identifiers to biological subfields.

Keywords: Diatoms, DNA barcoding, DNA barcode marker, Gene locus.

INTRODUCTION

Researchers have traditionally handled species identification and classification, providing a nomenclatural framework and a necessary prerequisite for a wide range of biological studies. Today's society must address a number of critical biological issues, including the importance of protecting natural, ensuring survival, preserving biodiversity, and preventing pandemics. To address these challenges, the 'DNA Barcode of Life' project seeks to create a standardised, rapid, and low-cost species identification method that is accessible to non-specialists (Frézal and Leblois 2008). With the development of PCR-based approaches for

species identification in the 1990s, the concept of a standardised mechanism of identifying molecules emerged gradually. Bacterial research, surveys of the diversity of microorganisms, and routine pathogenic strain diagnosis are the main applications of molecular identification to address the demand for identifying systems that are not culturally specific. The identification of eukaryotic pathogens and vectors, as well as food and forensic molecular identification, have all benefited from the widespread use of PCR-based techniques. Several universal molecular-based identification systems have been used for lower taxa but have not been successfully implemented for broader

scopes. The DNA barcode project's goal is not to create a molecular taxonomy tree, but rather to develop an easy new technique based on enormous biological data collected in the DNA barcode reference library. The DNA Barcode of Life data system allows for the collection, storage, analysis, and dissemination of DNA barcode records (Purty and Chatterjee 2016).

A DNA barcode is one or a few short gene sequences found in the genome that are distinct enough to identify species. By sequencing a very short standardised DNA sequence in a well-defined gene, DNA barcoding is a useful tool for taxonomic classification and species identification. Using this technique, complete species information can be obtained from a single specimen, regardless of morphological or life stage characteristics. The species is identified by using Polymerase Chain Reaction to amplify a highly variable region of the nuclear, chloroplast, or mitochondrial genome's DNA barcode region (Urbánková & Veselá, 2013). Nuclear DNA, chloroplast DNA, and mitochondrial DNA are some of the most commonly used regions for DNA barcoding. DNA barcodes can be used to group unknown species based on barcode sequences into previously known species or new species. The set of DNA barcode markers has been applied to specific taxonomic groups of organisms and has proven to be invaluable in understanding species boundaries, community ecology, functional trait evolution, trophic interactions, and biodiversity conservation. The use of NGS technology has increased the versatility of DNA barcodes across the 'Tree of Life,' habitats, and geographies, as new methodologies for characterising species are explored and developed (Purty and Chatterjee 2016). In an ideal world, a single gene sequence would be used to identify species across all taxa, from viruses to plants and animals. However, because the perfect gene has yet to be discovered, different barcode DNA sequences are used for animals, plants, microbes, and viruses.

Diatoms. A diatom is a photosynthetic, single-celled organism, which produces its own food in the same way as plants do. They are a major group of algae and one of the most common types of phytoplankton, joining the swarms of organisms that float on currents in the upper layers of the ocean and lakes (Ballesteros *et al.*, 2021). Diatoms can be found anywhere and everywhere. They can be found in rivers, oceans, lakes, bogs, damp rock surfaces, and even the skin of a whale. Diatoms are significant because they form the foundation of the food chain for both marine and freshwater microorganisms and animal larvae, and they are a major source of atmospheric oxygen, accounting for 20-30% of all carbon fixation on the planet. Diatoms can serve as environmental indicators of climate change and are used to make some household products such as pest/mite repellent and mild abrasives. Because diatoms have specific ecological requirements, they can also be used as environmental indicators,

informing us about what is going on in the environment. Diatom cell walls can be preserved in sediments for long periods of time, providing a record of past changes in lake systems (MacGillivray & Kaczmarzka 2011). Diatoms are the most common organisms in plankton and come in a wide range of shapes and sizes. Diatoms have silica cell walls, and each species has a unique pattern of tiny holes in the cell wall (frustule) through which they absorb nutrients and expel waste. When examined under a microscope, diatoms exhibit a wide range of shapes with numerous interesting and beautiful patterns (Liu *et al.*, 2020). Their shapes and structures are typically regular and symmetrical, and these characteristics are used to identify and classify them (Hamsher *et al.*, 2011). Phytoplankton are the smallest plankters, with sizes ranging from about 1mm to 7.5 micrometres, making them nearly invisible to the naked eye. All diatoms have a siliceous (glassy) exoskeleton composed of two halves that perfectly fit inside one another. Many diatoms remain as isolated cells and spend their entire lives adrift, whereas others form chains/clumps. Plankton samples were previously stored in formalin, which caused them to appear grey and lifeless – a stark contrast to their true colourful selves when fresh.

CRITERIA FOR IDENTIFICATION OF DIATOMS

Genes and gene locus. There have been numerous gene regions investigated for barcoding diatoms, and out of those genes, the mitochondrial cytochrome oxidase I gene (*cox1*), 18S nuclear rRNA, plastidial *rbcL* (ribulose-1,5-bisphosphate carboxylase oxygenase gene) and, nuclear rDNA ITS region has been widely used. Following several studies, it was discovered that the *rbcL* gene is less dynamic than the *cox1* gene within-species sampling, whereas it has been proven to be a favourable barcode marker in certain organisms.

The highly conserved 18S nuclear rRNA gene region has been used for environmental sample analysis and for phylogenetic research. It has been observed that the 18S rRNA has a high resolving and amplification power. *Cox1* having high polymerization has revealed the molecular inventories that differed the greatest from the expected inventories which are owing to the limited amount of reference barcodes created by Sanger sequencing. This low number is due to the fact of primer specificity. Because of the extremely varied v4 region and a large number of reference barcodes, 18S (including the v4 region) demonstrated a high degree of similarity between molecular and anticipated inventories. *rbcL* has demonstrated a higher polymorphism than 18S with a similar number of reference barcodes therefore the molecular inventories closest to expected inventories were obtained with *rbcL* (Zimmermann *et al.*, 2011).

DNA Barcoding. DNA barcode is a gene segment basically used in the species identification (Ács *et al.*,

2016). It has been growing very fast in recent years and becoming an important tool for biodiversity research and monitoring, as well as molecular phylogeny and evolution. The most widely used method for species identification and biological sample consistency is DNA barcoding. It can identify specimens to the genetic level. Fig. 1 illustrates the process of DNA Barcoding. When compared to traditional identification methods, DNA barcoding is more cost-effective, and it can even be used when just a little amount of sample is

available. After a trustworthy reference database has been created, the fundamental advantage of DNA barcoding is that it does not require specialised taxonomic expertise to identify particular samples. Additionally, since there is no need for reproductive material, identification can be carried out using small tissue samples from almost any part of the organism. It is also typically quick and repeatable. The lack of a single universal DNA region that can be used to all taxonomic groups is a disadvantage of the approach.

Diatom DNA Barcoding

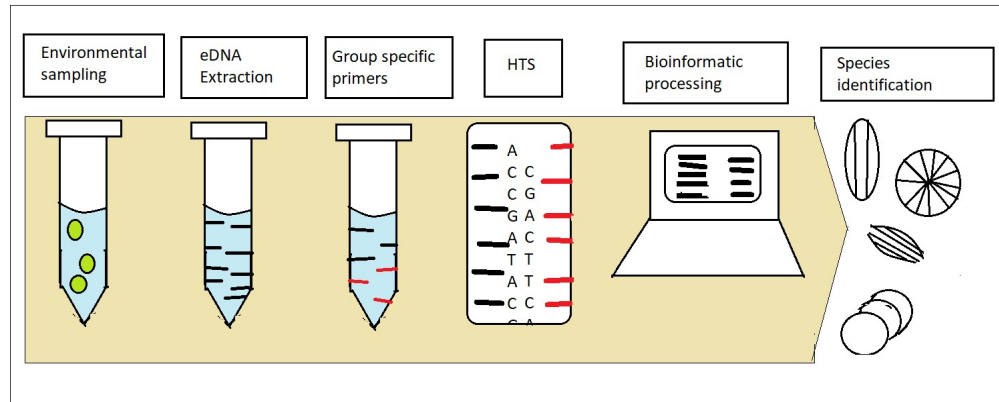


Fig. 1. Process of DNA barcoding.

DNA Barcode Marker. There are three basic requirements for a suitable barcode marker. They are (1) it should have an introductory sequence that can be easily amplified and sequenced in one single read (2) should be accompanied by a consensus sequence where universal primers can be inserted, and (3) should have the capability of resolving organisms at the genetic level. A suitable barcode marker can be determined by two conditions. Those two conditions are discriminatory power and universality. Discriminatory power states the marker's ability to distinguish between genetic diversity and universality refers to research problems such as the utilisation of primer pairs, the standard of sequences acquired, and the homology modelling challenges (Nauer *et al.*, 2022). The best functioning barcode markers for diatoms are currently available as follows: (i) the 3' end of the large subunit of the *rbcL* (*rbcL*-3 P), (ii) a 540 bp fragment situated 417 bp downstream of the start codon of the *rbcL* (540 bprbcL), (iii) the 5' end of the mitochondrial cytochrome c oxidase I gene (COI-5 P), (iv) a partial sequence of the large ribosomal subunit (D1-D3 LSU, usually either D1-D2 or D2-D3), and (v) the V4 sub-region of the small ribosomal subunit (V4 SSU) (Evans *et al.*, 2007). The 5.8 S gene, when paired with the second internal transcribed spacer, could be used as a diatom barcode marker which is having sufficient universality and good discrimination power. It has been rejected in many studies due to a lot of intraclonal variation, which made it difficult to link even closely related lineages. As a result of not meeting the

universality condition, all major subunits of the *rbcL* sequence are inappropriate for DNA barcoding. Due to its limited discriminatory power, the universal plastid amplicon was also proposed as a marker for all eukaryotic algae and cyanobacteria (Stiawan *et al.*, 2022).

ADVANTAGES OF DNA BARCODING

Documenting, phylogenetic revision, and the possibilities of using a microscope for identification. Barcodes will immediately aid taxonomy revision, enhancing the morphological data previously accessible, because they represent new information about organism genotypes. However, the new collections that barcoding development will necessitate and generate are arguably more important. Many additional specimens will have to be obtained and somatic mutation cultures isolated to create the reference barcodes. As a result of barcoding, vast new sources for diatom DNA barcoding and microscopic identification will emerge. Barcodes can also help to keep the nomenclature of living diatoms consistent (Zou *et al.*, 2021). Furthermore, because most types are permanently set in resin on slides and can only be studied by light microscopy, they frequently don't include enough information to limit the use of the defined term. Even when defragmented specimen is gathered for experimental results, determining how a name should be applied might be difficult. Barcoding will not eliminate these challenges immediately, but once barcodes are connected to type specimens and

made operational there would be a significantly less of a requirement to refer natural kind specimens.. Unlike the morphology of a physical object, a barcode sequence — essentially a molecular type which is clear and easily communicated (Mann *et al.*, 2010).

Species Discovery. Determination of a new species DNA barcoding was first created for classes of organisms like birds and fish that already had a thorough and accurate alpha taxonomy. The purpose of barcoding in such organisms, as well as in certain others where it has lagged due to methodological issues, is to easily identify them. In diatoms, however, there is still a significant amount of alpha taxonomy to be completed. As a result, many barcode sequences taken from wild populations or cultures would directly relate to nothing in the database even after having diatom species even after having collection of barcodes. Some represents previously undiscovered phenotypic variation and resembles a known species barcode while others are unidentified species that would need to be further characterized, described, and assigned to the proper supraspecific group (Rimet *et al.*, 2019).

So even though identification is the primary function of barcode but it can also be used for the evolutionary studies. It should be simple to align and involve both largely conserved and quickly evolving regions. One fair critique of barcode-based species finding is that it implies that speciation has not occurred below a certain level of divergence.

This is irrelevant since there is no causal connection between speciation and molecular divergence. When sister species are compared, neutral genetic variations build up in a pattern like a clock over time, although it may take a lengthy period following evolution by natural selection for sibling species to become reciprocally monophyletic about a barcode marker. The faster the barcode marker changes, the less likely newly developed species are to go unrecognized. Even if the two are linked, this is merely a modification of the concept of a molecular barrier for recognizing species, because compensatory base-change and speciation are not causally linked (Mann *et al.*, 2010).

New avenues for research into diatom biogeography and the biodiversity of living diatoms. Diatoms are generally dead when they are recognized due to the necessity to inspect minute details of frustule ornamentation and structure for specific identification, and it is not always obvious either they died as a result of the cleaning or if they were already killed when tested. As a result, determining either those specific groups were contemporaneous, coexisted in nature, or had crustal or allochronic origins is frequently challenging. At first glance, the method for detecting diatoms while they are still intact appears to be convincing. However, for the reasons mentioned, this is challenging to accomplish using microscopical techniques. So the first point is that there is a difference in refractive index between water and diatom silica and mountants like Naphrax is significantly less, frustule

features in living material are more difficult to perceive (Duarte *et al.*, 2020). The second reason includes that the cell wall patterning is hampered by chloroplasts and other cellular proteins, however, utilizing interference contrast optics and the use of elevated filters on photographic images can occasionally improve identification. Thirdly, chloroplast morphology gives extra relevant data because it hardly ever changes between most taxa, and even less so between centric diatoms, this benefit does not outweigh the loss of frustule detail (Mann *et al.*, 2010).

Limitations of Barcoding. The premise behind barcoding is that evolution is associative by a change in the barcode gene's sequence. As massive divergence of sequences is random rather instead of continuous. Even if the barcode's components rapidly increase, barcoding will fail to recognise certain lineages. Additional information will be required to identify such species. There is a further issue, which is caused by the 'weak' barcodes: some species may be impossible to barcode simply because they are largely undefined. As a result, barcoding has drawbacks and cannot identify all diatoms. Biological evolution, on the other hand, is a process in which various Species features emerge in a distinct order and at a different time in all characteristics, including morphology and reproductive isolation, are distinct lineages, may fail to differentiate species when utilized separately.

Some species or groups will almost certainly never be able to use the specified barcode. If *rbcL* were considered as an effective diatom barcode marker, barcoding would be impossible for several species that do not have a functional plastid and are facultatively anaerobic. Even DNA extraction appears to be challenging in some diatoms that produce a lot of mucilage, according to DNA barcoding for diatoms 567. It should also be noted that barcoding does not eliminate the need for microscopy. A lot would be lost if barcoding was seen as a substitute for microscopy rather than as an adjunct to it because many aspects of community structure and function, such as three-dimensional cell arrangement, motility, and cell-size spectra, cannot be determined without the use of an optical or microscopical technique.

Challenges in developing Barcoding for Diatoms. The principal challenges are (1) choosing the taxonomic basis for barcoding, (2) developing and testing candidate barcodes, and (3) generation of a sufficiently comprehensive set of barcodes to make barcode identification practical.

The phylogenetic framework of barcoding. One of the most difficult aspects of identifying diatoms is that several taxonomies are used. Diatom DNA barcoding is being held back by this dispute. The majority of researchers would most likely to create a sensitive enough barcode technology to distinguish including all of the new species they are or will be describing, including cryptic and pseudo-cryptic forms. As a result, taxonomists can select a molecular marker that

develops quickly, such as ITS-1 or -2, or COI. This is referred to as a 'strong' barcode. Those who have successfully used diatoms for bio monitoring and discovered that a crude taxonomy suffices for their needs, on the other hand, maybe a 'weak' barcode with little ability to discriminate. It may be manageable to create a barcode system that closely resembles the widely used freshwater flora, with morphological features replaced by molecular ones (Fei *et al.*, 2020). Understanding whether evolutionary change in diatoms is typically or always accompanied by differences in physical or chemical requirements, or specific to biotic factors, or whether clades of closely related species share the same niche, would be useful when deciding between "weak" and "strong" barcodes. To find a solution, speciation studies in depth, as well as ecological studies in diverse habitats and with various types of diatoms, are required. These studies are still in their early stages, but preliminary findings suggest that speciation is linked to niche, indicating that it is possible to improve the environmental monitoring resolution by using a 'strong' barcode. As a result, using a 'weak' barcode is likely to limit the exploitation of these organisms for bioengineering and biomonitoring, as well as genetic analysis, biodiversity, and ecology research. So this type of barcode has significant impacts. A strong barcode system's high resolution will almost certainly allow for future advancements in bio monitoring and ecological research. It will also enable the identification and study of cryptic species. Whereas a 'weak' barcode system tends to stabilise taxonomy (Kollár *et al.*, 2021). It would convert a classification based primarily on light microscopy, which is a molecular identification system, which is already recognised as inadequate for various research disciplines.

DEVELOPMENT AND TESTING OF CANDIDATE BARCODES

Up to this point, LSU rDNA, SSU rDNA, ITS rDNA, the universal plastid amplicon (UPA), *rbcL*, and COI have been tested. The criteria for evaluating barcodes are the same as for any other group of organisms: (A) universality, (B) practicability, and (C) discrimination. The term "power" refers to a marker's capacity for differentiation. The barcode's universality can be determined by putting it to the test on a phylogeny of the diatoms including wide range of taxa. There is no agreement based on the major diatom lineages' branching order but a universality test, however, must contain references from each of the major lineages of 'radial centric diatoms,' several 'multipolar centric diatoms,' and a diverse range of pennate diatoms. According to practicality, the barcode must be short enough to allow to reads in both directions with a fixed pair of primers, and analysis procedures must be simple. It does not necessitate the use of complex algorithms to achieve desired alignment. Practicality evolves and, as equipment and bioinformatics protocols

improve, becomes less of a constraint. Anyway, more tests are required, and the desired universality is most apparently found in incomplete LSU rDNA, incomplete ITS-1–5.8S–ITS-2, *rbcL* or selective *rbcL* and UPA. The subject utilising any of the rDNA regions is one of practicality: Intragenomic variation is common as a result numerous, non-identical rDNA cistron copies are formed and may be dispersed throughout one, two, or more loci.

Although there could be one dominant version, others would be plentiful enough to minimise precise reads during DNA amplification, considering the frequent length change due to insertions and deletions. Direct ITS sequencing is not possible in several genera of species. Furthermore, arrangement of rDNA sequences is incredibly hard, becoming more challenging with evolutionary detachment measured by functional constraints on molecule speciation. Alignment isn't required for identification because algorithms like BLAST can be used to compare sequences.

Moreover alignment ease of access becomes a key core challenge, given the current diatom taxonomy, if both species finding and classification are to be accomplished using barcode areas. In other rDNA regions, intragenomic variation may be insufficient and interspecific variation may be excessive to make them useful as barcodes. Protein-encoding genes, such as COI and *rbcL*, present serious fewer obvious issues that are produced than rDNA and can be easily combined and contrasted. Once the universality and practicability conditions are in place, the discrimination of barcode markers must be tested. Because of the existing system of taxonomy and the choice between 'weak' and 'strong' barcodes, this is the most difficult of the three factors to analyse in diatoms. Examining a barcode's performance in distinguishing between random selection of organisms with novel associations that can only be deduced from the barcode is meaningless. Evidently, only a few groups of species have been highlighted sufficiently to be used as diatom related species (Cristóbal *et al.*, 2020).

Creation of diatom barcodes and the emergence of a system of identification. Extracting material for producing the reference barcode is usually difficult in multicellular organisms: A single genotype is represented by a leaf or a scrap of tissue cut from any living organism, and enough DNA is provided for analysis and culture is necessary to supply diatoms cellular composition and DNA content. Though many diatoms have never been successfully cultured, others cannot be kept in culture indefinitely due to their mating system. As a result, diatoms are unrepresented and unbalanced in accumulation of cells, and efforts to isolate and propagate strains are renewed through barcoding diatoms (Kahlert *et al.*, 2021).

Sequencing is most likely the least difficult stage. It is imperative to preserve DNA whenever we need to check the barcodes and provide materials for upcoming research. The complete diatom marker has not yet

discovered but this is equally important as it may be still under development. Given the current state of sampling, in the future all specimens in the barcode database could have additional barcode markers if DNA is made available. Indeed, it would be incredibly difficult to redo the massive culturing and vouchering effort that will be required to establish the barcode database (Hiransuchalert *et al.*, 2022).

For taxa that are resistant to isolation and culture, it is now possible to amplify the entire genome from a single or a few cells. However, due to the difficulty in profiling and verifying the morphology of the cells extracted, this is problematic for barcoding. So there are the following steps:

— Consensus on two or more barcode areas (on the basis of their universality, practicability, and discriminatory power) that all species need to be sequenced. Presently, the most promising candidates are 3'-rbcL and partial LSU rDNA.

— All the data for the preparation of common protocols for culturing, vouchering, characterization, DNA preservation, and the use of primers, more specifically has been added to the central database.

— Further testing of potential barcode markers, as well as on-going attempts to determine model group's species limits. Existing markers and protocols are being continuously improved.

— Increasing the effort being put into culturing, vouchering, characterization, and identification. All these processes can be assigned to technical experts, but these are areas that require significant new funding. Initially, specific habitats and model groups should be prioritised, but there should also be a relatively broad coverage. Contributions from every taxonomists who study diatoms will be required to confirm that identification numbers for barcodes are properly related with the current taxonomy through a microscope, and approaches to overcome strong linkages to the extensive amount of alpha-taxonomic research and the barcode endeavour performed in diatoms should be sought (Smith *et al.*, 2022).

APPLICATIONS OF DNA BARCODING

DNA barcodes are used in a variety of fields, including taxonomy, ecology, biosecurity, and food safety. One of the DNA barcoding's primary goals is to accelerate the process of cataloging biodiversity through the use of standardized genetic markers for species identification. Molecular barcodes aid in the completion of the biodiversity inventory by 1) revealing cryptic diversity at different taxonomic levels 2) recognizing species in taxa with no distinguishing morphological features. DNA barcodes can also aid in the resolution of long-standing nomenclatural debates, resulting in the taxonomic revision of poorly defined morphospecies. In ecology and conservation biology, DNA barcoding is also widely used. Molecular barcodes are sometimes used to detect and monitor invasive and endangered

species by tracing their DNA contained in hair, faeces, or water samples (Al-Meshhdany & Hassan 2020).

The analysis of non-degraded DNA in stomach contents reveals specific species diets or interspecies interactions, such as the predation pressure of some invasive species. DNA barcodes are also commonly used for pest species detection and food quality control. Metabarcoding, also known as environmental DNA barcoding, is another emerging barcoding technique which uses genetic markers to identify individuals found in environmental materials such as dirt seawater etc (Naeem *et al.*, 2019). Short DNA barcodes are used in metabarcoding to classify species diversity or to detect specific species in environmental DNA extracts (Ahmed *et al.*, 2022).

The advancement of next-generation sequencing (NGS) technologies capable of producing millions of sequences at a low cost prompted the development of metabarcoding

(<https://www.biorxiv.org/content/10.1101/2022.05.04.490577v1>). The results of next-generation sequencing studies revealed a huge variety of aquatic eukaryotes, including many promising lineages and undiscovered species. Metabarcoding has also been used to measure the environmental impacts of human activities, and to monitor freshwater benthic diversity (Pawlowski & Holzmann 2014).

CONCLUSION

DNA barcoding is a system for rapid and accurate species identification that will improve access to the ecological system. It has many applications in various fields such as identification of new species, evolutionary relationships, biomonitoring and bioassessment, forensics, cryptic species, and databasing. DNA barcoding is a useful technique for identifying organisms at the molecular level. This technique includes polymerase chain reaction (PCR) to amplify a fragment of the gene, which is then sequenced and compared to a database of known organisms. The purpose of this study is to obtain the polymerase chain reaction (PCR) primers and reagents required for DNA barcoding on a wide range of taxonomic groups. This technology reduces the number of organisms that must be accumulated in the field while also decreasing the time between collection and identification.

FUTURE SCOPE

When compared to the conventional approaches for species identification, DNA barcoding is more effective and requires less skill. DNA barcoding would help to determine species diversity using complicated and ecological species. It can change as technology advances. It can aid in the discovery of global biodiversity. It will also streamline bio monitoring. Biological context will also be provided for ancient collections.

Acknowledgement. Authors are thankful to Lovely Professional University for all the opportunities provided for the work.

Conflict of Interest. None.

REFERENCES

- Ács, É., Ari, E., Duleba, M., Dressler, M., Genkal, S. I., & Kiss, K. T. (2016). Pantocsekiella a new centric diatom genus based on morphological and genetic studies. *Fottea*, 16(1), 56-78.
- Ahmed, S., Ibrahim, M., Nantasenammat, C., Nisar, M. F., Malik, A. A., Waheed, R., & Alam, M. K. (2022). Pragmatic Applications and Universality of DNA Barcoding for Substantial Organisms at Species Level: A Review to Explore a Way Forward. *BioMed Research International*, 2022.
- Al-Meshhdany, W. Y., & Hassan, F. M. (2020). Five Diatom Species Identified by using Potential Application of Next Generation DNA Sequencing. *Bulletin of the Iraq Natural History Museum*, 16(1), 39-61.
- Ballesteros, I., Terán, P., Guamán-Burneo, C., González, N., Cruz, A., & Castillejo, P. (2021). DNA barcoding approach to characterize microalgae isolated from freshwater systems in Ecuador. *Neotropical Biodiversity*, 7(1), 170-183.
- Cristóbal, G., Blanco, S., & Bueno, G. (2020). Modern Trends in Diatom Identification. Springer International Publishing.
- Duarte, S., Vieira, P. E., & Costa, F. O. (2020). Assessment of species gaps in DNA barcode libraries of non-indigenous species (NIS) occurring in European coastal regions. *Metabarcoding and Metagenomics*, 4, e55162.
- Fei, C., Zou, S., Wang, T., Wang, C., Kemuma, N. D., He, M., & Wang, C. (2020). A quick method for obtaining high-quality DNA barcodes without DNA extraction in microalgae. *Journal of Applied Phycology*, 32(2), 1165-1175.
- Frézal, L., & Leblois, R. (2008). Four years of DNA barcoding: current advances and prospects. *Infection, Genetics and Evolution*, 8(5), 727-736.
- Hamsher, S. E., Evans, K. M., Mann, D. G., Pouličková, A., & Saunders, G. W. (2011). Barcoding diatoms: exploring alternatives to COI-5P. *Protist*, 162(3), 405-422.
- Hiransuchalart, R., Kutako, M., Watchasit, S., Kaewduang, M., Setthamongkol, P., Chindudsadeegul, P., & Jaritkhuan, S. (2022). Morphological and molecular comparison of three centric diatoms (Bacillariophyceae: Chaetoceros) isolated from the coast of Eastern Thailand. Research Square. DOI: <https://doi.org/10.21203/rs.3.rs-1406725/v1>
- Kahlert, M., Baitel, B., Chonova, T., Karjalainen, S. M., Schneider, S. C., & Tapolczai, K. (2021). Same same, but different: The response of diatoms to environmental gradients in Fennoscandian streams and lakes—barcodes, traits and microscope data compared. *Ecological Indicators*, 130, 108088.
- Kollár, J., Pinseel, E., Vyverman, W., & Poulickova, A. (2021). A time-calibrated multi-gene phylogeny provides insights into the evolution, taxonomy and DNA barcoding of the Pinnulariagibba group (Bacillariophyta). *Fottea*, 21(1), 62-72.
- Liu, M., Zhao, Y., Sun, Y., Li, Y., Wu, P., Zhou, S., & Ren, L. (2020). Comparative study on diatom morphology and molecular identification in drowning cases. *Forensic science international*, 317, 110552.
- MacGillivray, M. L., & Kaczmarek, I. (2011). Survey of the efficacy of a short fragment of the rbc L gene as a supplemental DNA barcode for diatoms. *Journal of Eukaryotic Microbiology*, 58(6), 529-536.
- Mann, D. G., Sato, S., Trobajo, R., Vanormelingen, P., & Souffreau, C. (2010). DNA barcoding for species identification and discovery in diatoms. *Cryptogamie Algologie*.
- Nacem, M., Ali, J., Hassan, M. Z., Arshad, B., Rao, M. H. I., Sarmad, M. S. K., ... & Hussain, M. U. (2019). Novel Approach towards DNA Barcoding as a Tool in Molecular Biology and Biological Activities of Cyclotides with Particular Emphasizes at Molecular Level. *Biological Forum -An International Journal*, 11(2): 83-96.
- Nauer, F., Marchi, F., Naves, M., Plastino, E. M., & Oliveira, M. C. (2022). A species dilemma: utilizing biological species concept to reconcile DNA barcode and morphological divergences for *Hypnea pseudomusciformis* (Gigartinales, Rhodophyta). *Journal of Applied Phycology*, 34(1), 697-708.
- Pawlowski, J., & Holzmann, M. (2014). A plea for DNA barcoding of foraminifera. *The Journal of Foraminiferal Research*, 44(1), 62-67.
- Purty, R. S., & Chatterjee, S. (2016). DNA barcoding: an effective technique in molecular taxonomy. *Austin J Biotechnol Bioeng*, 3(1), 1059.
- Rimet, F., Gusev, E., Kahlert, M., Kelly, M. G., Kulikovskiy, M., Maltsev, Y., ... & Bouchez, A. (2019). Diat. barcode, an open-access curated barcode library for diatoms. *Scientific Reports*, 9(1), 1-12.
- Smith, S. A., Santoferrara, L. F., Katz, L. A., & McManus, G. B. (2022). Genome architecture used to supplement species delineation in two cryptic marine ciliates. *Molecular Ecology Resources*.
- Stiawan, E., Rachmawati, Y., Nurachman, Z., & Sutomo, S. (2022). Identification of Tropical Marine Diatom of *Chaetoceros* sp. from Bokor Island through SEM Imaging and Genetic Marker Analysis: *Indonesian Journal of Chemical Studies*, 1(1), 1-7.
- Urbánková, P., & Veselá, J. (2013). DNA-barcoding: A case study in the diatom genus *Frustulia* (Bacillariophyceae). *Nova Hedwigia*, 142, 147-162.
- Zimmermann, J., Jahn, R., & Gemeinholzer, B. (2011). Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organisms Diversity & Evolution*, 11(3), 173-192.
- Zou, S., Bao, Y., Wu, X., & Wang, C. (2021). DNA Barcoding Diatoms From China With Multiple Genes. *Frontiers in Marine Science*.

How to cite this article: Suvechha Kabiraj, Suman Jyoti Bhuyan and Umesh Goutam (2022). Species Identification in Diatoms using DNA Barcoding: An Overview. *Biological Forum – An International Journal*, 14(3): 979-985.